YALE

N65-85777

<u>Status Report on NASA Grant Ns-G-208-62</u>

This report covers the period May 1, 1964 to May 30, 1965. The report summarizes the present status, both experimental and theoretical of minimal self replicating entities.

The smallest free living organisms approach sizes at which the structure of matter itself imposes limits on the complexity and function of the cells. The linear dimensions of the most minute free living forms are the order of magnitude of one thousand times the linear dimensions of atoms. For this reason a study of this branch of biology brings sharply into focus certain limiting problems about the process of self duplication in biology. In addition, all organelles in cells of this size range must of necessity be in the general size range of macromolecules or reasonably small aggregates of macromolecules.

In this report on the minimum self replicating unit we will focus attention on organisms which are capable of replication in the absence of a host cell. This is done to avoid the difficulty of knowing exactly how much of the host apparatus to assign as being an essential part of the replication cycle. Thus, the class <u>Microtatobiotes</u> will not be regarded as self replicating in the sense of being antonomously self replicating. However, where it is instructive we will present comparable data on viruses.

The minimum self replicating system will then be regarded as the minimum unit capable of forming a clone from a single organism in the absence of any other living cells. Minimum implies an attempt at simplicity, a concept not always easy to define. Two views are possible at this point, the smallest free living organism in terms of mass or the organism with the smallest genome. At the present, one species fills both these specifications. Table 1 and Figure 1 give data on cell mass and genome size for a variety of minute organisms.

The first part of this report will consist of a description of the observed size and properties of the smallest free living cells that we have been able to investigate. It must be noted that the experimental search for the smallest cell is always open ended and limited by our techniques of searching particularly in regard to the choice of growth medium and growth conditions. Following the discussion of the experimental material we will set up a model of the minimum cell consistent with a current viewpoint of molecular biology. Thirdly, we will briefly examine some of the theoretical aspects relating to very small cells.

Since the smallest free living cells that have been investigated are pleuropneumonia-like organisms* (PPLO, *Mycoplasma*) we will begin with a general description of this family. The problems of PPLO taxonomy have been discussed by Freunt (1958) in his monograph "The *Mycoplasmataceae*." The family designation is *Mycoplasmataceae* of which one genus *Mycoplasma* has been recognized. PPLO are minute cells which lie in the general size range between the bacteria and the viruses. Because of pleomorphism their size overlaps both other groups of microorganisms. PPLO are capable of autonomous self replication and a single entity will give rise to a colony when grown on an appropriate nutrient agar surface. They are distinguished from the bacteria in three principal ways:

1. PPLO lack a rigid cell wall and are bounded only by a unit membrane (Klieneberger-Nobel, 1962).
2. Viable units will be found in the filtrate through a cellulose acetate filter

---

* A number of minute bacteria have been reported particularly in genera *Veillonella*, *Dialister*, and *Achromobacter*. From the available evidence it appears that these are larger than the small PPLO.

of 0.3μ pore diameter (Morowitz, Tourtellotte and Pollack, 1963). Most strains have viable units which pass through a 0.22μ pore diameter filter. Since the cells are not rigid, filter pore diameter does not demonstrate that the average cell diameter if less than 0.22μ.

3. All strains except one contains cholesterol (Klieneberger-Nobel, 1962). In addition, there appear to be no major biochemical differences between these cells and other cellular systems. The pathways, intermediates, and macromolecules appear to be the same as in other cellular systems.

Since the studies on the properties of PPLO have been carried out on many strains, attention will be concentrated on three of these which will prove to be of special interest: H39, Mycoplasma laidlawii and Mycoplasma gallisepticum.

H39 is a strain initially isolated from human tissue. Among the various strains tested in our laboratory it is the smallest by the following criteria:

1. Genome size
    a. Measured by electron microscopy of the displayed DNA from cells lysed on a protein film.
    b. Measured by DNA per clone forming unit.
2. Cell mass based on DNA, RNA and protein per clone forming unit.
3. Cell size from electron microscopy.
4. Filterability - H39 yields the highest relative titer through a 0.22μ pore size filter.

A typical post division cell of this strain has the following properties:

1. A non aqueous mass of the order of 10^{-14} gms. This corresponds to less than 10^9 atoms.
2. A genome of molecular weight 350×10^6.
3. Approximately 400 ribosomes.

4. The protein mass which is the order of 5×10^{-15} gms and is the equivalent of about 60,000 protein molecules of 50,000 molecular weight.
5. At a pH of 7 there are an average of two hydrogen ions per cell.

It should be noted that a genome of 350×10^6 molecular weight units has a relatively limited coding capacity. This molecular weight corresponds to 5.5×10^5 base pairs or 1.1×10^6 bits of information. Looked at in a somewhat different way, the total protein message that could be encoded in this genome is 1.8×10^5 amino acids. If we assume an average of 450 amino acids per protein then 400 types of proteins could be encoded in this genome. The functional capacity of H39 cells is limited by the size of the genome. By comparison, the smallest bacterial genome that has been reported is 800×10^6 for H. influenzae.

The studies on the ultrastructure of this strain are in a preliminary stage. The cells are ellipsoids with average linear dimensions the order of 0.4μ. The sizes may vary between 0.2μ and 1.0μ. The cell is bounded by a unit membrane and the ribosomes appear to be mainly in a zone around the periphery of the cell. Replication appears to proceed by binary fission. The two daughter cells may be of unequal size.

The strain Mycoplasma gallisepticum is larger than H39. It has, however, yielded considerable information on detailed morphology as well as the nature of the replication cycle. For this reason we will review these features (Maniloff, Morowitz and Barnett, 1965a;1965b; Maniloff, 1965). It should be noted, however, that no evidence exists at the moment that the particular morphological features of this strain are of general occurrence in the pleuropneumonia-like organisms.

There is considerable size variation in this strain. A typical post division cell is shown negatively stained in Figure 2. It is approximately 0.5 x 0.8μ in size and shows three well defined areas: a bleb about 1000 Å in diameter, an infrableb region, and a large cell body. In Figure 3 these same general features are shown in thin section micrographs of the same strain. The bleb is highly structured. The main body of the cells shows the ribosomes lining the interior of the cell membrane with the DNA in a central nuclear region. Replication proceeds in the following way. In the region opposite to the bleb a new bleb infrableb structure begins to grow. The cell elongates and the DNA replicates. The ribosomes continue to line the interior of the cell membrane in the central body of the cell. The cell further elongates, the nuclear material pulls apart and ribosomes fill the intranuclear space. The cell then pulls apart and divides. Figure 4 shows a dividing cell.

A number of features appear from the microscopic description of this strain.

1. The highly structured bleb which is clearly related to the replication process appears to be a new organelle, not previously observed in microorganisms.
2. Cells of this strain are totally ordered. There appears to be no subvolume of the cell that can be regarded as a solution phase. There is evidence that the ribosomes themselves exist in orderly arrays (Maniloff, Morowitz, and Barnett, 1965a).
3. The replication process is programmed in a rather precise way with respect to the subcellular morphology.
4. Cells of a rather limited genome showed a very elaborate subcellular morphology.

The third strain of interest is Mycoplasma laidlawii B. It is of special interest here because of two types of study that have been performed: a) growth on a defined medium and b) characterization of membrane subunits.

M. laidlawii will grow on a defined medium made up of about 40 constituents. The genome size has not been measured with great precision but probably lies between 550 and 900 x 10^6. The ability to grow on a defined medium is the ultimate test of autonomy, and is significant to find a pleuropneumonia-like organism that meets this criteria.

When highly purified membrane preparations of M. laidlawii are treated with sodium lauryl sulfate (SLS) they dissolve giving rise to subunits which sediment as homogeneous 3.4s particles. The sedimentation of SLS membrane particles have been studied in H_2O and D_2O. By using solvents of different density, information is obtained on the partial specific volume of the sedimenting particle. From these values it appears that the subunits are lipoproteins. When the dissolved material is dialyzed to remove the SLS and add divalent metal ions, structures are reconstituted which show the same structure in thin section electron microscopy as the initial membrane preparations (Razin, Morowitz, and Terry, 1965). The reformed material consists of "unit membranes" about 70 Å in thickness. Thus, the membrane appears to be structured of rather homogeneous subunits. We have subsequently found a number of membrane systems which exhibit the behavior just described for M. laidlawii preparations. That is, they disaggregate into lipoprotein subunits which can then be reaggregated into unit membrane structures.

Having briefly reviewed the pertinent properties of the pleuropneumonia-like organisms we may turn our attention to postulating the minimum biological self replicating system consistent with present day knowledge of molecular biology. We begin by listing the empirical generalizations which will define the system we wish to discuss.

1. The minimal system will be a cell. Morphologically, this consists of structures surrounded by a "unit membrane" approximately 75 Å in thickness.
2. The system will be capable of autonomous self replication. In the absence of a host cell or other living entity the minimal cell will be able to extract appropriate molecules from its environment and gives rise to two or more entities similar to the original so that at least two of these are capable of themselves being self replicating.
3. Information storage is in double stranded DNA. Thus, between 500,000 and

1,000,000 molecular weight of DNA must be allocated for each enzymatic function.

4. Protein synthesis proceeds through activated amino acids and transfer RNA. Synthesis takes place at a ribosome-messenger RNA site.
5. Nucleic acid synthesis is a direct template readout involving polymerases.
6. A minimum of energy metabolism must be maintained to supply ATP from ADP and phosphate.
7. The cell is biochemically normal; that is, the building blocks are the usual amino acids, nucleotides and sugars which are ubiquitous in biological systems. In addition, the cell is approximately 75% water.

The next step is to postulate the number of enzymatic functions necessary for replication. In protein synthesis we need twenty activating enzymes and at least one polymerizing enzyme. For nucleic acid synthesis we need two polymerizing enzymes plus another group involved in the synthesis of nucleotide triphosphates. We will assume that nine enzymes are involved in these functions. A simplified energy metabolism can operate with about eight enzymes. The synthesis of the lipid moiety of the membrane lipoproteins requires another group of enzymes and a minimum number of five will be postulated. This totals 45 and assuming 750,000 molecular weight of DNA per enzyme, the genome to code these functions must be 33,750,000 molecular weight units. In addition, the following must be encoded: ribosomal protein, ribosomal RNA, transfer RNA, and membrane protein. These will bring the minimal genome molecular weight to about 40,000,000. In addition, if the cell has two molecules of each enzyme type, two molecules of each type of transfer RNA, 1,000,000 molecular weight of messenger RNA and three 70s ribosomes, then we must add an additional 16,000,000 molecular weight bringing the total to 48,000,000 or 8×10^{-17} gms. This mass, as an 25% aqueous suspension, would occupy a sphere 425 Å radius. If we now enclose this sphere in a unit membrane the resulting cell is about 1000 Å in diameter. This is the smallest hypothetical cell that we can envision within the context of current biochemical thinking. It is almost certainly a lower limit, since we have allowed no control functions, no vitamin metabolism and extremely limited intermediary metabolism. Such a cell would be very vulnerable to environmental fluctuation.

In a population of H39 cells there appears to be viable cells with an average diameter of less than 3000 Å. Since the minimum hypothetical cell has a diameter of over 1000 Å there is a limited gap in which to seek smaller cells.

In light of the present discussion of the minimum cell and previous considerations of some of the orderly principles of biomolecular structure it is pertinent to raise questions about the possibility of synthesizing these cells. We will first make the simplifying assumption that we need only to synthesize the appropriate molecular structure and it will then function as a cell. This conclusion has been amply confirmed by experiments in cryobiology which demonstrate that cells will not lose thier self replicating property after being exposed to temperature near absolute zero. Since only structure persists at these temperatures it is apparent that specification of structure is sufficient to specify the minimum self replicating unit. We conclude that in principle the synthesis of a living cell requires a skilled organic chemist, someone capable of assembling the appropriate atoms and connecting them with the appropriate bonds.

The next point to be raised is the problem of how difficult it would be to synthesize a structure of 10^9 atoms. Here information theory lends some insight at an idealized level. Using optimal binary coding and specifying each atom it would take about 7×10^9 bits to specify such a structure. A set of instructions of this length could approximately be encoded in a set of volumes with a total message length somewhat longer than that of the Encyclopedia Britanica.

What is made clear by the general principles of biomolecular structure is that this figure is a gross overestimate. For example, if a cell has ten thousand identical membrane subunits it is only necessary to encode the information once, whereas in the previous estimate it was encoded ten thousand times. In addition, it is clear that it is not necessary to localize soluble molecules in the cell nor is it necessary to localize all organelles with very great precision. It may also be noted that if we operate with natural subunits rather than atoms the complexity will be considerably reduced. Thus, if the previous calculation started with amino acids, bases, sugars, fatty acids, etc., instead of atoms the information would be reduced to 5.3×10^8 bits.

An alternative approach is to calculate the information in the genome and assume that this must be synthesized in four forms for assembly into the cell. The four forms are DNA language, RNA language, protein language and substrate language. Such an analysis leads to an information estimate of 4×10^6 bits for a cell of the complexity of H39. An instruction of this information capacity could clearly be encoded in a single volume. The largest structure that can presently be synthesized have an information content of 2×10^3 bits. Thus we are quite far from synthesizing anything of the complexity of a free living cell. Nevertheless, there is nothing to suggest that, from an informational point of view, there are any barriers to this type of synthesis.

These considerations lead to a rather simple view of the functional complexity of minimal cellular systems. We assume that primary information storage is in the DNA and that utilization of 400 chemical steps is sufficient to guarantee cell replication in a sufficient number of cases. We further assume that specification of macromolecular structure is almost sufficient to specify organelle structure and that the structure and organization of organelles will provide an adequate basis for cell replication. The elucidation of the synthesis of organelles from macromolecules will then provide the background for the detailed study of the replication procdss.

The existence of very minute free living forms brings into focus a number of biological questions about replication and control in living systems. The smaller the cells, the more sharply are the questions posed. It seems reasonable to believe that the detailed characterization of these small systems will begin to provide answers to a number of these questions.

Table 1

Cell Radius (Microns)	Non-aqueous Cell Mass (gms)	Number of Atoms in Non-aqueous Cell Mass	Total Number of Macromolecules Molecular Weight 50,000	Fraction of Cell Volume in Membrane (percent)	Structures in This Size Range
.30	30.5×10^{-15}	22.9×10^{8}	361,000	9.73	P. pneumonae
.28	24.8×10^{-15}	18.6×10^{8}	298,000	10.3	M. laidlawii
.26	19.8×10^{-15}	14.9×10^{8}	238,000	11.2	H. influenzae
.24	15.7×10^{-15}	11.8×10^{8}	189,000	12.1	M. gallisepticum
.22	12.0×10^{-15}	9.00×10^{8}	144,000	13.0	V. parvulus
.20	9.02×10^{-15}	6.77×10^{8}	109,000	14.1	D. pneumosintes
.18	6.59×10^{-15}	4.94×10^{8}	79,400	16.0	H39
.16	4.62×10^{-15}	3.47×10^{8}	55,600	17.5	
.14	3.08×10^{-15}	2.31×10^{8}	37,000	19.4	
.12	1.95×10^{-15}	1.46×10^{8}	23,400	23.1	Vaccinia virus
.10	1.13×10^{-15}	8.47×10^{7}	13,600	27.0	
.08	5.78×10^{-16}	4.34×10^{7}	6,960	33.2	
.06	2.44×10^{-16}	1.83×10^{7}	2,930	41.1	Influenza virus

Bibliography

Freunt, E. A., 1958. The Mycoplasmataceae, Statens Serum Institute, Copenhagen.

Klieneberger-Nobel, E. (1962). Pleuropneumonia-like Organisms (PPLO) Mycoplasmataceae. Academic Press, N.Y.

Maniloff, J., Morowitz, H. J., and Barnett, R. J. (1965a). J. Cell Biol., 25, 139.

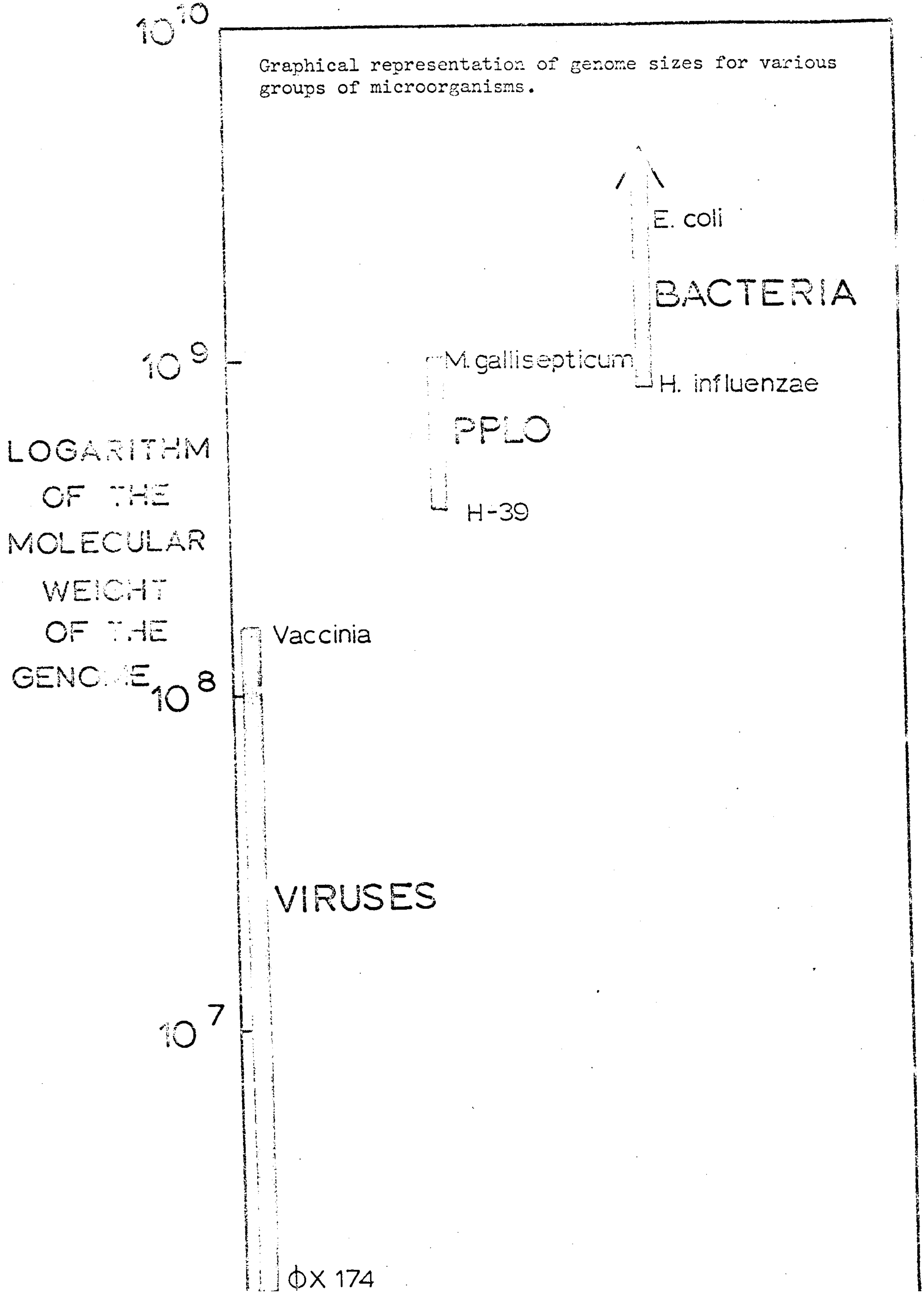
Maniloff, J., Morowitz, H. J., and Barnett, R. J. (1965b). J. Bacteriol., in press.

Maniloff, J. (1965). Ph.D. Thesis, Yale University.

Morowitz, H. J., Tourtellotte, M. E., and Pollack, M. E. (1953). J. Bacteriol., 85, 134-136.

Razin, S., Morowitz, H. J., and Terry, T., in press.

Figure 1



Graphical representation of genome sizes for various groups of microorganisms.
10^10
10^9
10^8
10^7
LOGARITHM OF THE MOLECULAR WEIGHT OF THE GENOME
E. coli
BACTERIA
H. influenzae
M. gallisepticum
PPLO
H-39
Vaccinia
VIRUSES
ΦX 174

Figure 2

Single cell of *Mycoplasma gallisepticum*, fixed in glutaraldehyde and negatively stained.

Figure 3

Single cell of _Mycoplasma gallisepticum_ fixed in glutaraldehyde, epon embedded, thin sectioned and stained with uranyl acetate. (72,800x)

Figure 4

Dividing cell of <u>Mycoplasma gallisepticum</u> prepared as in Figure 3. (72,800x)

Appendix I

Considerable progress has been made on the theoretical studies on irreversible statistical mechanics and evolution. This material is currently being put in more organized form for the summer colloquium in theoretical biology. It will be covered in more detail in a subsequent report.